資料WG21－1－2

「偽・誤情報に対するコンテンツモデレーション等の在り方」に関する主な論点（案）

２０２４年５月２２日

デジタル空間における情報流通の健全性確保の在り方に関する検討会
ワーキンググループ事務局

※　本資料は、ワーキンググループにおける議論のたたき台として、主査の指示の下、事務局にて論点となり得る事項を幅広に列挙したものであり、今後、実際の議論状況等を踏まえ、記載内容や構成等が変更される可能性がある。

今後のWGにおける検討の進め方（案）

資料WG18−1−2より抜粋

- 資料WG16-1-1「デジタル空間における情報流通の健全性に関するWG検討課題（案）」に記載の各課題について、第16回会合において、検討時間に限りがある中で優先順位をつけながら議論を進めるべきとのご意見をいただいたところ、効率的に検討を進める観点から、全体を大きく

1. 情報流通の健全性を巡る課題一般

2. （その中でも特に）広告収入を基盤としたビジネスモデルに起因する課題

の２つに分類した上で、それぞれの課題への対応の在り方について、既存の法令等による対応が可能な部分とそれ以外の部分の切り分けを意識しつつ、下記の順序で検討していくことについて、どう考えるか。

No.	大分類	小分類
①	1. 情報流通の健全性を巡る課題一般への対応の在り方	a. 災害発生時等における情報流通の健全性確保の在り方
②		b. マルチステークホルダーによる連携・協力の在り方
③		c. 偽・誤情報に対するコンテンツモデレーション等の在り方
④		d. 情報伝送PFが与える情報流通の健全性への影響の軽減に向けた方策の在り方
⑤	2. 広告収入を基盤としたビジネスモデルに起因する課題への対応の在り方	a. 広告の質の確保を通じた情報流通の健全性確保の在り方
⑥		b. 質の高いメディアへの広告配信に資する取組を通じた情報流通の健全性確保の在り方
⑦		c. 情報伝送PFによる発信者への経済的インセンティブ付与や収益化抑止の在り方
⑧		d. 情報流通の健全性確保の観点から見たレコメンデーションやターゲティングの在り方

論点１：対応を検討すべき「偽・誤情報」の定義・範囲

➢ 情報伝送PFは、デジタル空間における情報流通の主要な場となっており、その中で偽・誤情報が流通・拡散すること等により、個人の意思決定の自律性への影響や、権利侵害、社会的混乱その他のフィジカル空間への影響が発生・増幅し得るところ、こうした影響の軽減等に向けて対応を検討すべき「偽・誤情報」の範囲をどのように考えることが適当か。<☞**参考資料WG21-1-2 pp.1～8参照**>

- 前提として、**「偽・誤情報」をどのように定義するか**。
  ※海外では、**発信者の主観的意図**に着目し、誤りが含まれる情報のうち、発信者が事実でない事項を事実であると誤認・誤解させる**意図を持って発信**した情報を「偽情報」（disinformation）、そのような**意図を持たずに発信**した情報を「誤情報」（misinformation）と定義する事例があるが、どうか。

- 上記定義に該当する「偽・誤情報」のうち、**社会的な影響の軽減に向けて対応の検討が必要な範囲**をどのように考えるか。例えば次のような要素に着目することが考えられるが、どうか。

① **違法性・社会的影響の重大性**
※当該情報そのものが有する**違法性・権利侵害性**があるか。
※当該情報の**客観的な有害性**や、それが流通することによる**社会的影響の重大性・明白性**があるか。
例）救急・救命活動への影響、健康被害、株価への影響、公共インフラの損壊、詐欺被害、風評被害
※諸外国では、市民の健康・安全等に害を及ぼし得ることを制度的対応が必要な「偽情報」「誤情報」の要件に含める事例があるが、どうか。
パロディ・風刺など、重大な影響を及ぼすおそれの小さい情報を制度的対応が必要な「偽情報」「誤情報」から除外する事例があるが、どうか。

② **検証可能性・容易性**
※「誤りが含まれる情報」であることについての**検証可能性・容易性（明白性）**があるか。
参考）LINEヤフー株式会社（2024年2月22日・WG第3回会合（検討会第9回会合））
「政府機関・ファクトチェック機関など信頼できる機関によるファクトチェック結果に基づき明らかな偽・誤情報と判断されるものについて対応」
「プラットフォーム事業者においては、各種の情報・時間的制約から何が「偽情報」であるか範囲を確定することが困難な場合も」

③ **その他**
※「誤りが含まれる情報」のみならず、**誤解を招く（ミスリーディングな）情報**をどう捉えるか。
例）必ずしも誤りは含まれていないが、文脈上誤解を招く情報
※「内容」に誤りが含まれている情報のみならず、なりすましアカウントによる投稿など、発信者の**「名義」に誤りが含まれる情報**をどう捉えるか。

論点２：偽・誤情報の流通・拡散を抑止するための「コンテンツモデレーション」の類型

➢ 論点１で検討した範囲の「偽・誤情報」に対し、情報伝送PFがその流通・拡散を抑止するために講ずる措置（いわゆるコンテンツモデレーション）※として、どのようなものが有効と考えられるか。<☞**参考資料WG21-1-2 p.8参照**>

※信頼できる情報の受信可能性の向上（いわゆるプロミネンス）を通じて間接的に偽・誤情報の拡散を抑止する措置を含む。

コンテンツモデレーションの主な類型			概　要
	抑止効果	可視性への影響	
①発信者に対する警告表示	低？	影響なし	不適切な内容を投稿しようとしている、又は直近で投稿したことが判明している旨の警告を表示する措置（投稿自体は可能）
②収益化の停止	中？	影響なし	広告を非表示にしたり、広告報酬の支払いを停止することにより、収益化の機会を失わせる措置
③ラベルの付与	中？	影響なし	情報発信者の信頼性等を見分けるためのラベルを付与する措置（本人確認を行っていない利用者の明示等）
	中？	一部影響あり	情報の信頼性等を見分けるためのラベルを付与する措置（ファクトチェック結果の付与等）
④表示順位の低下	高？	一部影響あり	投稿されたコンテンツを、受信者側のおすすめ欄等の表示候補から外したり、上位に表示されないようにする措置
⑤情報の削除	高？	影響あり（可視性ゼロ）	投稿された情報の全部又は一部を削除する措置（新規投稿等は可能）
⑥サービス提供の停止・終了、アカウント停止・削除	高？	影響あり（可視性ゼロ）	サービスの一部から強制退会、又はその一部の利用を強制終了し、新規投稿等をできないようにする措置 アカウントの一時停止又は永久停止（削除）を実施する措置

⑦信頼できる情報の受信可能性の向上（いわゆるプロミネンス）

論点３：偽・誤情報に対するコンテンツモデレーションの実施の促進方策（総論）

➢ 論点１で検討した範囲の「偽・誤情報」に対し、情報伝送PFがコンテンツモデレーションを実施することを促進等するための方策として、どのようなものが必要かつ適当か。

- 例えば次のような方策が考えられるが、どうか。

（１）対応の透明性の確保を通じた過不足ない実施の確保

（例）

① コンテンツモデレーションに関する**基準や手続を事前に策定・公表**

② コンテンツモデレーションの実施要否等の判断に関与する**人員等の体制に関する情報を公表**

③ 上記①の基準の**運用状況を事後に公表**

④ コンテンツモデレーションを実施した場合に、**その旨及び理由を発信者に通知**

（２）対応の迅速化を通じた実施の促進

（例）

① 外部からの**コンテンツモデレーション申請窓口を整備・公表**

② 上記①の窓口を通じて申請があった場合に、**一定期間内にコンテンツモデレーションの実施の要否・内容を判断し、申請者に判断結果を通知**

③ コンテンツモデレーションの実施の要否・内容を判断するための**体制を整備**

④ 一定の条件※の下で行ったコンテンツモデレーションにより発信者が被った損害について、**情報伝送PFを免責**
※例えば、行政機関等の特定の第三者からの要請を受けてコンテンツモデレーションを実施した場合など

論点３：偽・誤情報に対するコンテンツモデレーションの実施の促進方策（総論）

（つづき）

（３）可視性に影響しない措置の確実な実施

・「収益化の停止」や「発信者に関するラベルの付与」など、情報そのものへの可視性に影響しないコンテンツモデレーション（又はそれ以上の措置）を体制を整備して確実に実施

（４）可視性に影響する対応も含む措置の確実な実施

・「情報の削除」や「アカウント停止等」など、可視性に影響するコンテンツモデレーションも含め、体制を整備して確実に実施

（５）上記（１）～（４）の組合せによる対応

- 上記のような方策の**実効性を制度的に担保する必要性**について、どう考えるか。制度的な対応を行わない場合、**どのような対応があり得るか**。

※例えば、上記（１）～（５）のような対応を情報伝送PFに義務付けることが考えられるが、どうか。

※上記（４）を制度的に担保する措置（コンテンツモデレーションの類型のうち、「情報の削除」や「アカウント停止等」の義務付け）については、過度な情報削除やアカウント停止が行われるおそれがあることや、発信者の表現の自由に対する実質的な制約をもたらすおそれがあること等から慎重であるべきとの考え方があり得るが、どうか。

※①権利侵害情報に該当する偽・誤情報、②違法情報に該当する偽・誤情報、③その他の論点１で検討した範囲の偽・誤情報など、偽・誤情報の特性・性質に応じた対応を考えるべきか。表現の自由の確保等との関係でどのように考えれば良いか。

論点４：いわゆる「インプレッション稼ぎ」への対応

※スライド１の⑦の論点関係

➢ 論点３に関
を付与する
拡散をはじめ
対応するため

資的インセンティブ
偽情報の発信・
ろ、こうした状況に

- 例えば、情
が、どうか。
- その際、「

① 情報伝
定の効果が
見込ま

② 情報自

③ 過度な

- 上記のよう
を行わない場合、
どのような

☞ **資料２１－１－３参照**

論点５：偽・誤情報に対するコンテンツモデレーションの実施の契機

➢ 論点３で検討した方策について、どのような契機でコンテンツモデレーションを実施することが適当か。

他人の権利を侵害する違法な偽・誤情報 や 行政法規に抵触する偽・誤情報の場合

- 例えば、次の主体からの申出・要請を契機としてコンテンツモデレーションを実施することが考えられるが、どうか。
  - ① **自己の権利を侵害されたとする者（被害者）**
  - ② 行政法規を所管する**行政機関**（その委託や認証を受けた機関を含む。）

- 上記②の場合、**行政機関による恣意的な申出・要請を防止**し、**透明性・アカウンタビリティを確保**するとともに、**過度な申出・要請に対し発信者や情報伝送PFを救済**するための方策として、どのようなものが適当か。
  例えば、次のような方策が考えられるが、どうか。
  - ① 行政機関において、**申出・要請に関する手続等※を事前に策定・公表**　※事後救済手段を含む
  - ② 行政機関において、**実際に行った申出・要請の状況を事後的に公表**
  - ③ 申出・要請に応じて実施されたコンテンツモデレーションにより発信者が被った損害について、**情報伝送PFを免責**
  - ④ コンテンツモデレーションを実施した情報伝送PFにおいて、**行政機関の名称等の情報を発信者に通知**

違法ではない偽・誤情報の場合

- 例えば、次のような主体からの申出・要請を契機としてコンテンツモデレーションを実施することが考えられるが、どうか。他の方法もあり得るか。こうしたプロセスを構築する場合、どのような点に留意点が必要か。
  - ① 当該情報付近に広告を表示された**広告主**
  - ② **ファクトチェック機関**
  - ③ その他情報伝送PFが自らあらかじめ定めて公表した**信頼できる第三者**

論点６：コンテンツモデレーションに関する透明性・アカウンタビリティの確保

➢ 論点３で検討した方策について、情報伝送PFによるコンテンツモデレーションが過不足なく実施されていることについて、利用者を含む社会一般が確認し、情報伝送PFのサービスに対する信頼性を向上させるための方策として、どのようなものが必要かつ適当か。

- 例えば次のような措置の実施を情報伝送PFに求めることが考えられるが、どうか。

① コンテンツモデレーションに関する**基準や手続を事前に策定・公表**

② コンテンツモデレーションの実施要否等の判断に関与する**人員等の体制に関する情報を公表**

③ 上記①の基準の**運用状況を事後に公表**

④ コンテンツモデレーションを実施した場合に、**その旨及び理由を発信者に通知**

論点７：偽・誤情報の発信を抑止するための方策

- 以上のほか、論点１で検討した範囲の「偽・誤情報」の発信を抑止するための方策として、どのようなものが考えられるか。

  - 例えば、情報伝送PFが発信者に対し、次のような方策を実施することが考えられるが、どうか。

    ① **アカウント登録時の本人確認の厳格化**

    ② **botアカウントの抑止策の導入（アカウントの有料化等）**

  - 上記のような方策の**実効性を制度的に担保する必要性**について、どう考えるか。制度的な対応を行わない場合、**どのような対応があり得るか**。

論点８：偽・誤情報への対応策の実施を求める情報伝送PFの範囲

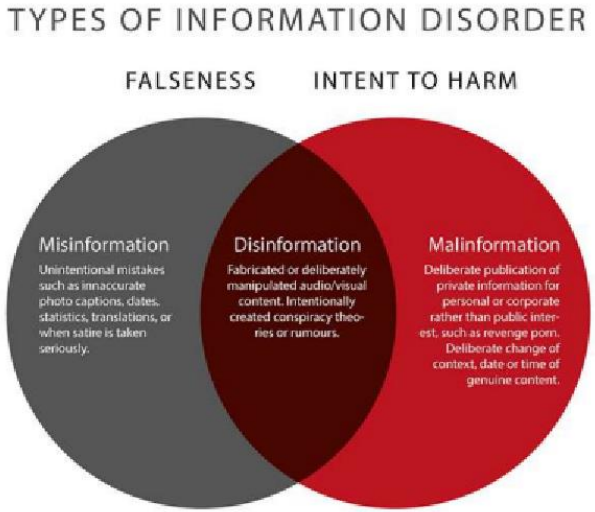
- 以上で検討した偽・誤情報の流通・拡散や発信への対応策の実施を、どの範囲の情報伝送PFに求めるか。
  - 例えば、偽・誤情報の流通の頻度や社会に与える影響の深刻度という観点から、**利用者数**や、**サービスの目的・性質**などを勘案し、一定の要件を満たす**大規模な情報伝送PF**のみを対象とすることが考えられるが、どうか。

【参考１】諸外国を中心とした「偽・誤情報」等の定義の例

EU①

◆ 欧州評議会レポート「Information Disorder」（2017年9月）での定義

- **偽情報**（Dis-information）：
  虚偽の情報であって、個人、社会集団、組織又は国家を害する目的で意図的に生成されたもの。
- **誤情報**（Mis-information）：
  虚偽であるが、害を生じさせる意図をもって生成されたものではない情報。
- **悪情報**（Mal-information）：
  事実に基づく情報であって、個人、組織又は国家に害を与える目的で利用されるもの。



（出典）Council of Europe report DGI(2017)09, *Information Disorder: Toward an interdisciplinary framework for research and policy making,* Sep. 27, 2017
（https://www.coe.int/en/web/freedom-expression/information-disorder）

◆ 欧州委員会コミュニケーション（2018年4月）での定義

- **偽情報**（disinformation）は、検証可能な、虚偽又は誤解を招く情報で、経済的利益を得るため又は公共を欺くことを目的として生成、表示、拡散され、それによって公共への損害が生じ得るものとして理解されている。
- **公共への損害**は、民主的な政治プロセス及び政策形成プロセスや、EU市民の健康、環境又は安全の保護等の公益に対する脅威から成る。
- 偽情報は、誤報、風刺及びパロディ、又は明白に確認されている党派性の強いニュース及び解説を含まない。

（出典）Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Tackling online disinformation: a European Approach*, Apr. 26, 2018
（https://digital-strategy.ec.europa.eu/en/library/communication-tackling-online-disinformation-european-approach）

EU②

◆ 欧州民主主義行動計画（2020年3月）での定義

- **誤情報**（misinformation）：
  有害な意図を持たずに共有されながらも、その効果は未だ有害である虚偽の、又は誤解を招くコンテンツ。
  例）虚偽の情報を善意で友人や家族に共有する場合

- **偽情報**（disinformation）：
  欺き、又は経済的若しくは政治的利得を確保する意図を持って拡散され、公共への損害を生じさせ得る虚偽の、又は誤解を招くコンテンツ。

- **情報影響操作**（information influence operation）：
  偽情報と組み合わせて独立した情報源を抑圧することを含む幅広い欺罔的手段を用いて対象となる聴衆に影響を与えるために国内又は国外の主体によって行われる組織的な試み。

- **情報空間における外国による干渉**（foreign interference in the information space）：
  個人の政治的意図の自由な形成及び表現を妨げるために外国の国家主体又はその代理主体によって行われる抑圧的・欺罔的な試み。

（出典）Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European democracy action plan, Mar. 12, 2020
(https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan_en#documents)

⇨ **偽情報に関する強化された行動規範（the Strengthened Code of Practice on Disinformation 2022）でも踏襲（※）**

※同行動規範はさらに、「偽情報」が以下を含まないことを明記：
- 誤解を招く広告
- 誤報
- 風刺及びパロディ
- 明白に確認されている党派性の強いニュース及び解説

豪州①

◆ 偽情報及び誤情報に関する豪州行動規範（Australian Code of Practice on Disinformation and Misinformation）での定義

- **偽情報**（Disinformation）：次の３要件を全て満たすデジタルコンテンツ（※1）。
  - a. 検証可能な形で誤っている、誤解を招く、又は詐欺的である。
  - b. 欺罔的行動（※2）を通じてデジタルプラットフォームの利用者間で伝播される。
  - c. その拡散が損害（※3）を生じさせる合理的な可能性がある。
- **誤情報**（Misinformation）：次の３要件を全て満たす（多くの場合は合法な）デジタルコンテンツ（※1）。
  - a. 検証可能な形で誤っている、誤解を招く、又は詐欺的である。
  - b. デジタルプラットフォームの利用者間で伝播される。
  - c. その拡散が損害（※3）を生じさせる合理的な可能性がある（が、明確に意図されてはいないかもしれない）。

※1 デジタルコンテンツ（Digital Content）：本行動規範の署名者が所有し、又は運営するプラットフォーム上でオンライン配信される豪州の利用者を対象とするコンテンツであって、AIアルゴリズムの利用を通じるなどの自動的手段によって人為的作成され、操作され又は修正されたコンテンツを含む。

※2 欺罔的行動（Inauthentic Behaviour）：スパム及びその他の形態の詐欺的、操作的又は大量で攻撃的な（自動的システムを通じて実行される場合もある）行動が含まれ、また利用者のオンライン上の会話に人為的な影響を与えること及び／又はデジタルプラットフォームの利用者にデジタルコンテンツを伝播するよう促すことを意図した行動が含まれる。

※3 損害（Harm）：以下のいずれかに対する確度が高く深刻な損害。
- 不正投票、投票妨害、誤った投票情報など、民主的政治・政策決定プロセス
- 市民の健康保護、社会で周辺化された人々若しくは社会的弱者の保護、公共の安全及び治安又は環境などの公共財

注）「確度が高く深刻な損害」には合理的に予測できない損害を含まない。

豪州②

◆ 2023年通信法改正案（Combating Misinformation and Disinformation Bill）での定義

- デジタルサービス上の**誤情報**（misinformation）：次の４要件を全て満たすコンテンツ（※1）の拡散。
  a. 誤っている、誤解を招く、又は詐欺的な情報を含んでいる。
  b. 適用除外（※2）に当たらない。
  c. 豪州内の1人又はそれ以上のエンドユーザーにデジタルサービス上で提供される。
  d. デジタルサービス上でのその提供が重大な損害を生じさせ、又はこれに寄与する合理的な可能性がある。
- デジタルサービス上の**偽情報**（disinformation）：次の５要件を全て満たすコンテンツ（※1）の拡散。
  a. 誤っている、誤解を招く、又は詐欺的な情報を含んでいる。
  b. 適用除外（※2）に当たらない。
  c. 豪州内の1人又はそれ以上のエンドユーザーにデジタルサービス上で提供される。
  d. デジタルサービス上でのその提供が重大な損害を生じさせ、又はこれに寄与する合理的な可能性がある。
  e. 拡散し、又は拡散させている人物が、当該コンテンツをもって他人を欺罔する意図を有している。

  注）偽情報は、外国勢力による、又は外国勢力に代わって行われる偽情報（の拡散）を含む。

※1 コンテンツ（content）：文書、データ、発話・音楽その他の音声、画像（動画を含む）その他の形式による、又はこれらを組み合わせた形式によるコンテンツ。

※2 適用除外に当たるコンテンツ：
- 真に娯楽、パロディ又は風刺の目的で作成されたコンテンツ
- 専門的なニュースコンテンツ
- 国内外の認定済み教育機関により、又は認定済み教育機関のために作成されたコンテンツ
- 政府・自治体により認証されたコンテンツ

ニュージーランド

◆ オンライン上の安全及び損害に関するニュージーランド行動規範（Aotearoa New Zealand Code of Practice for Online Safety and Harms）での定義

- **偽情報**（Disinformation）：次の３要件を全て満たすデジタルコンテンツ。
  a. 検証可能な形で誤っている、誤解を招く、又は詐欺的である。
  b. 欺罔的行動を通じてデジタルプラットフォームの利用者間で伝播される。
  c. その拡散が損害（※）を生じさせる合理的な可能性がある。
- **誤情報**（Misinformation）：次の３要件を全て満たす（多くの場合は合法な）デジタルコンテンツ。
  a. 検証可能な形で誤っている、誤解を招く、又は詐欺的である。
  b. デジタルプラットフォームの利用者間で伝播される。
  c. その拡散が損害（※）を生じさせる合理的な可能性がある（が、明確に意図されてはいないかもしれない）。

※ 損害（Harm）：利用者の安全及び／又はデジタル情報エコシステムの完全性に対して差し迫った深刻な脅威をもたらし、現実世界での危害につながる可能性のある行為者、行動及び／又はオンライン上のコンテンツ。具体的には以下の7つのテーマのいずれかに該当するもの。

① 児童の性的搾取及び虐待
② ネットいじめ又はハラスメント
③ ヘイトスピーチ
④ 暴力の扇動
⑤ 暴力的又はグラフィックなコンテンツ
⑥ 誤情報
⑦ 偽情報

EU・豪州・ニュージーランドの比較①

偽情報の定義

豪州とNZはほぼ同じ定義を採用している。EUには不正な行為による伝播という要素がなく、風刺やパロディ等の除外がある。

小分類	EU	豪州	NZ
偽情報	偽情報とは、人を欺いたり、経済的・政治的利益を確保したりする意図で流布される虚偽または誤解を招く内容であり、公衆に害を及ぼす可能性がある。 「偽情報」という概念には、誤解を招く広告、報道の誤り、**風刺やパロディ、明らかに党派的なニュースや論評は含まれず**、拘束力のある法的義務、自主規制の広告規範、誤解を招く広告に関する基準を損なうものではない。”	3.2.このコードが焦点を当てている偽情報の側面は、次のとおりである。 A. 検証可能な虚偽、誤解を招く、または欺瞞的なデジタルコンテンツ; B. 不正な行為(Inauthentic behaviours)を通じてデジタルプラットフォームのユーザー間で伝播されている。そして C. その流布が害を引き起こす合理的な可能性がある。	（i） 検証可能な虚偽、誤解を招く、または欺瞞的なデジタルコンテンツ; （ii） 不正な行為(Inauthentic behaviours)によってデジタルプラットフォームのユーザー間で伝播される;そして （iii） その流布が害（harm）を及ぼす合理的な可能性がある ※不正な行為は定義規定なし ※害とは、ユーザーの安全および/またはデジタル情報エコシステムの完全性に差し迫った深刻な脅威をもたらし、現実世界での危害につながる可能性のある行為者、行動、および/またはオンライン上のコンテンツを指す。

【出典】本WG第8回会合（2024年3月18日）配付資料WG8-1「EU・豪州・ニュージーランド・英国における行動規範の策定状況」（株式会社野村総合研究所ご発表資料）

EU・豪州・ニュージーランドの比較②

誤情報の定義

虚偽・誤解を招く表現、ユーザーによる伝播、害を起こす意図がない点は３者共通している。豪州・NZには害をもらたす合理的可能性の要素が加わり、豪州では風刺等の除外がある。

小分類	EU	豪州	NZ
誤情報	誤情報とは、有害な意図なしに共有される虚偽の、あるいは誤解を招くような内容のことであるが、例えば、人々が善意で友人や家族と虚偽の情報を共有する場合、その影響は依然として有害でありうる。 “誤情報とは、悪意なく共有された虚偽または誤解を招くコンテンツだが、その影響は依然として有害である可能性がある。(例:人々が善意で友人や家族と虚偽の情報を共有した場合）”	3.6.誤情報とは： A. 検証可能な虚偽、誤解を招く、または欺瞞的なデジタルコンテンツ （多くの場合は合法）; B. デジタルプラットフォームのユーザーによって伝播される;そして C. その**流布が危害を引き起こす合理的な可能性**がある （しかし、明確に意図されていないかもしれない） 。 4.4.誤情報ではないコンテンツ： 以下のコンテンツは、本規範に基づく誤情報ではない。 A. **娯楽 （風刺やパロディを含む） または教育目標のために誠実に作られたコンテンツ**; B. オーストラリア州または連邦政府によって許可されたコンテンツ; C. 第5.23条から第5.25条に従うことを条件として、政治広告またはオーストラリアの法律に基づいて登録された政党によって許可されたコンテンツ; D. 専門的なニュースコンテンツ。 本セクション4.4のAからDに該当するコンテンツは、不正な行為によって伝播された場合、偽情報の定義に該当する可能性がある。	（i） 検証可能な虚偽、誤解を招く、または欺瞞的なデジタルコンテンツ （多くの場合は合法）; （ii） デジタルプラットフォームのユーザーによって伝播される;そして （iii） （**合理的に可能性が高いが、明確に意図されていない可能性**がある）害をもたらす ※害とは、ユーザーの安全および/またはデジタル情報エコシステムの完全性に差し迫った深刻な脅威をもたらし、現実世界での危害につながる可能性のある行為者、行動、および/またはオンライン上のコンテンツを指す。

【出典】本WG第8回会合（2024年3月18日）配付資料WG8-1「EU・豪州・ニュージーランド・英国における行動規範の策定状況」（株式会社野村総合研究所ご発表資料）

英国

オンライン安全法における偽誤情報の位置づけ

偽誤情報の定義

■ オンライン安全法における違法コンテンツの定義の中に、偽情報は含まれていない。

- 参考：英国政府は、「偽情報（disinformation）を、人々に危害を与えるため、あるいは政治的、個人的、金銭的利益を得るために、人々を欺き、誤解させることを意図した虚偽の情報および／または操作された情報を意図的に作成し、広めること」と定義している。また、「誤情報（misinformation）とは、不注意による虚偽の情報の拡散である」と定義している

オンライン安全法における偽誤情報に関連する項目

■ 同法の中で、**偽誤情報に関連す項目としては、大きくは三つ**あり、偽誤情報のアドバイザリー委員会の設置（後頁参照）と、新たな虚偽通信罪の規定、OFCOMのメディアリテラシー義務に関する規定である。

■ 虚偽通信罪については、同法の179条で規定されており、虚偽であると知っている情報を、情報が心理的または身体的危害を与えることを意図していた場合、および、その情報を送信することについて合理的な理由がない場合に違反となるとしている。

■ OFCOMのメディアリテラシー義務については、同法の165条において、規制対象サービスを利用する際に、自分自身や他人を守ることができる方法について、一般市民の認識と理解を高めるための措置を講じることをOFCOMに求めており、例示として「偽情報と誤報の性質と影響」を理解することが挙げられている。

オンライン安全法における偽誤情報の位置づけに対する評価

■ 英国のファクトチェック団体であるFull Factは、オンライン安全法は利用規約にどのような内容を盛り込み、どのように対処・監督するかの規定が不足しているとし、誤情報の拡散を防ぐために十分ではないとの意見を表明している。

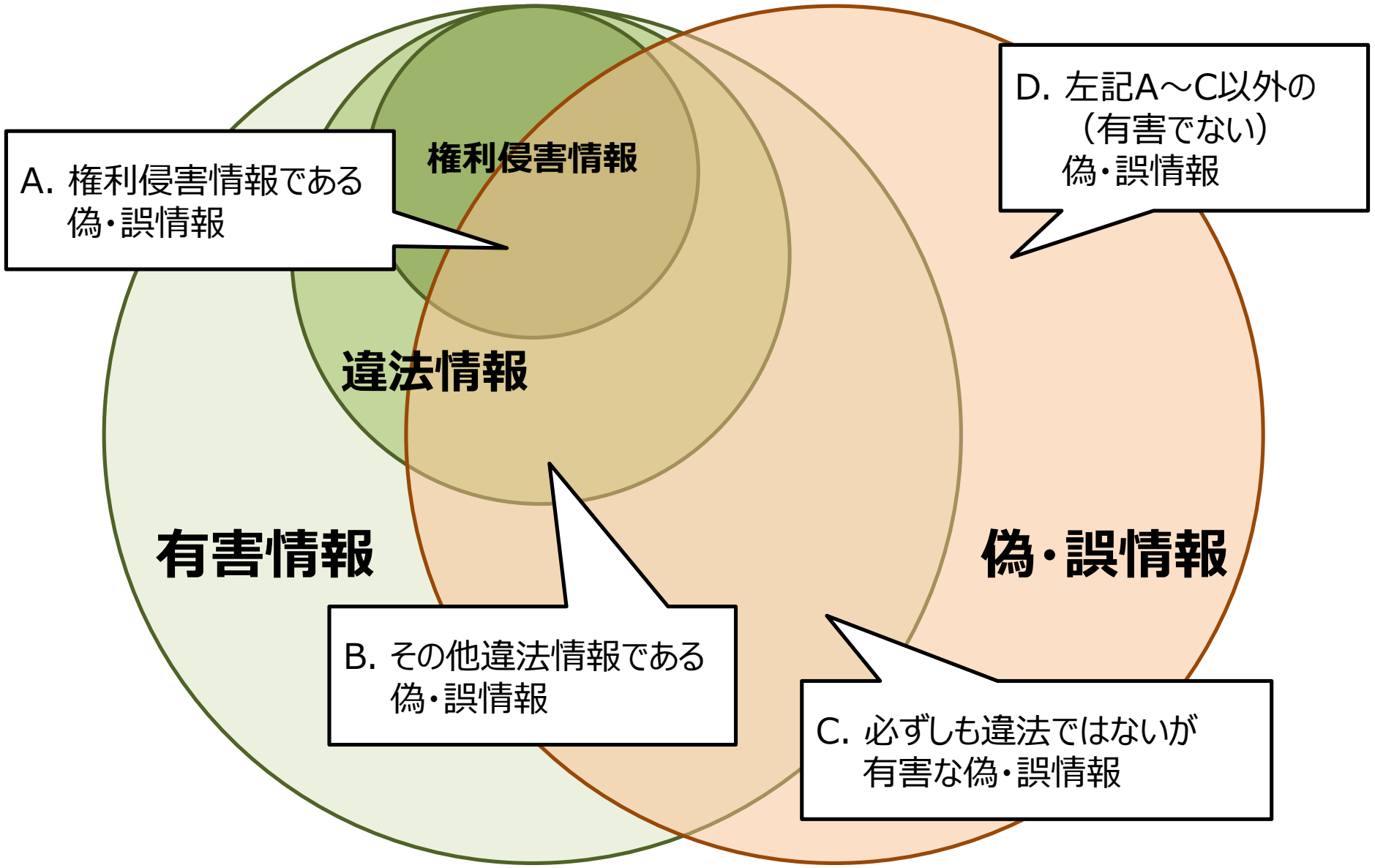
https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0003/
https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted
https://www.OFCOM.org.uk/__data/assets/pdf_file/0027/211986/understanding-online-false-information-uk.pdf

「違法情報」・「有害情報」・「権利侵害情報」の概念と「偽・誤情報」の関係①

◆「インターネット上の違法・有害情報への対応に関する研究会」最終報告書（2006年8月）

- 「**違法な情報**」とは、法令に違反したり、他人の権利（法律上保護される利益を含む。以下同じ。）を侵害したりする情報を・・・いうものとする。
- 「**有害な情報**」とは、違法な情報ではないが、公共の安全や秩序に対する危険を生じさせるおそれのある情報や特定の者にとって有害と受け止められる情報をいうものとする。
- インターネット上を流通する情報に対するプロバイダや電子掲示板の管理者等による対応については、平成14年5月に施行された・・・プロバイダ責任制限法・・・において、インターネット上の情報の流通により他人の権利が侵害されている場合にプロバイダや電子掲示板の管理者等が行う対応によって生じ得る損害賠償責任の範囲が規定されている。さらに、同法の実務的な運用指針として、プロバイダ責任制限法ガイドライン等検討協議会において、「名誉毀損・プライバシー関係ガイドライン」、「著作権関係ガイドライン」、「商標権関係ガイドライン」（以下、併せて「関係ガイドライン」という。）がそれぞれ策定され、プロバイダや電子掲示板の管理者等は、関係ガイドライン等を参考にして、流通により他人の権利を侵害する情報（以下「**権利侵害情報**」という。）への対応を行ってきたところである。
- ところが、インターネット上には、**権利侵害情報以外の違法な情報**（わいせつ情報、違法薬物の販売広告情報等の法令に違反する情報。以下「社会的法益等を侵害する違法な情報」という。）、**違法な情報ではないが公共の安全や秩序に対する危険を生じさせるおそれのある情報**（爆発物の製造方法に関する情報、人を自殺に誘引する情報等）や**特定の者にとって有害と受け止められる情報**（違法ではないアダルト情報等）等が流通しているところ、これらの情報については、プロバイダ責任制限法及び関係ガイドラインが適用されるものではないため、プロバイダや電子掲示板の管理者等が情報について送信防止措置等の対応を行った場合における法的責任や、特定の情報の流通が法令に違反するか否か等の判断に関する指針が存在しない状況である。

「違法情報」・「有害情報」・「権利侵害情報」の概念と「偽・誤情報」の関係②



【参考２】情報伝送PFによるコンテンツモデレーションの実施の促進等に関連する既存の制度の例（EU）

「違法コンテンツ」・「利用規約に違反する情報」への対応

5. DSAの対象となる情報（1/2）

- 「違法コンテンツ」（illegal content）について定義（第3条）されており、仲介サービス事業者による対応の対象とされている。
- 「違法コンテンツ」について、それらの削除を直接義務づける規定はないが、違法コンテンツについては、司法・行政機関からの措置命令への対応結果の報告義務（第9条）や、明白な違法コンテンツを頻繁に投稿する利用者に対するサービス提供停止義務（第23条）がある。
- 「利用規約に違反する情報」（information incompatible with terms and conditions）については、利用規約の内容に関して規定（第14条第1項）されており、利用規約に違反する情報に対する対応として、透明性報告義務（第15条等）や発信者への対応理由の通知（第17条）、苦情処理・紛争解決（第20条・第21条）等を通じた透明化された対応が求められている。

項目	DSAでの定義	DSAでの言及箇所（抜粋）
違法コンテンツ（illegal content）	• **第3条にて定義されている** “「違法なコンテンツ」とは、それ自体または製品の販売やサービスの提供を含む活動に関連して、EU法またはEU法に準拠している加盟国の法律に準拠していない情報を意味する。”（第3条(h)）	• 前文12項：「違法コンテンツ」の概念には、違法なコンテンツ、製品、サービス、活動に関連する情報をカバーするために広く定義されるべきとされている ✓ 具体例として、それ自体が違法となるヘイトスピーチ、テロリストのコンテンツ、差別的コンテンツや、適用される規則が違法行為に関連するという事実を考慮して違法とする情報を指す、**児童の性的虐待を描写した画像の共有、同意のない違法な私的画像の共有、オンラインストーカー行為、非準拠または偽造品の販売、消費者保護法に違反する製品の販売またはサービスの提供、著作権で保護された素材の非正規使用、違法な宿泊サービスの提供、生きた動物の違法な販売など**が挙げられている。 • 第9条：**違法コンテンツに関する司法・行政機関からの措置命令への対応結果を報告すること**とされている • 第23条：事前に警告を発した上で、**明らかな違法コンテンツを頻繁に提供するサービスの受信者に対して、合理的な期間、サービスの提供を停止しなければならない**とされている ✓ 前文63項：明らかに違法なコンテンツを頻繁に提供したり、明らかに根拠のない通知や苦情を、それぞれ本規則に基づき設置されたメカニズムや制度の下で頻繁に提出したりすることによるオンラインプラットフォームの悪用は、信頼を損ない、関係者の権利や正当な利益を害する。したがって、このような悪用に対して、適切で、比例的で、効果的なセーフガードを設ける必要がある。
利用規約に違反する情報（information incompatible with terms and conditions）	• 利用規約に含めるべき内容は、第14条で規定されている • 違法コンテンツまたは利用規約に違反する情報に対する対応については、透明化することが複数の条項で定められている	• 第14条：利用規約の中に、提供されるコンテンツについて、仲介サービスの利用にあたって利用者に課す条件についての情報を含めるものとされている • 第15条：透明性の報告義務の中で、報告される情報は違法コンテンツや利用規約に違反した情報等によって分類されることとされている • 第17条：利用規約に違反する情報に対して削除等の対応をした場合、利用規約に違反するとみなした理由について通知しなければならないとされている • 第20条：違法コンテンツまたは利用規約に違反する情報であることを理由として行った対応に対する内部苦情処理システムへのアクセスを提供しなければならないとされている ✓ 前文58項：サービスの受信者は、オンラインプラットフォームのプロバイダーによる、コンテンツの違法性や、利用規約との不適合に関する特定の決定に対して、容易かつ効果的に異議を申し立てることができるべきである

出所）EU-Lex「Document 32022R2065」https://eur-lex.europa.eu/eli/reg/2022/2065/oj

「偽情報」への対応

5. DSAの対象となる情報（2/2）

■「偽情報」(disinformation)については、「違法コンテンツ」の定義には含まれていないが、違法コンテンツと併記される形で前文に複数の項目で記載されており、社会的な悪影響を与えるリスクとして偽情報が明記されている。

項目	DSAでの定義	DSAでの言及箇所（抜粋）
偽情報 （disinformation）	• 具体的な定義はされていないが、前文104項で偽情報の説明がある "**偽情報**や操作的な悪用行為、未成年者への悪影響など、システミックリスクが社会と民主主義に及ぼしうる負の影響も考慮すべき領域である。 これには、意図的に不正確な、あるいは**誤解を招くような情報**、または経済的利益を得る目的で作成されたボットや偽アカウントを使用するなど、**偽情報を含む情報の増幅**を目的とした協調的な操作が含まれ、これらは特に未成年者などサービスの受け手である弱者にとって有害である。"（前文104項）	• 前文2項：**社会的リスクとして、違法コンテンツとオンライン上の偽情報を併記** • 前文9項：**DSAの目的**として、**オンラインでの違法コンテンツの流布と、偽情報やその他のコンテンツの流布が引き起こす可能性のある社会的リスクへの対処**を明記 • 前文69項：広告ターゲティングによる悪影響として、偽情報キャンペーンへの加担の可能性を明記 • 前文83項：VLOPやVLOSEのシステミックリスクが公衆衛生に関連する組織的な偽情報キャンペーンからも発生する可能性があることを明記 • 前文84項：VLOPやVLOSEの提供者はシステミックリスクを評価する際、違法ではないが特定されたシステムリスクに寄与する偽情報などの、誤解や欺瞞的なコンテンツを増幅するためのどのようにサービスが利用されるかについて、特に注意を払うべきと明記 • 前文88項：システミックリスクが偽情報キャンペーンに関連する場合には、VLOPやVLOSEの提供者は意識向上の活動も検討する必要があることを明記 • 前文95項：オンライン広告のもたらす可能性のあるリスクの例として、違法な広告等と併記して、偽情報を明記 • 前文104項：**偽情報に対して、違法コンテンツとは別に、自主規制によって検討されるべき特定の分野であると明記** • 前文106項：**DSAがEUで確立されている自主規制の基礎になりうるとした上で、その具体例として偽情報に関する行動規範について言及**

【出典】本WG第5回会合（2024年3月4日）配付資料WG5-1-6「EU DSA法（Digital Services Act）の概観」（株式会社野村総合研究所ご発表資料）

コンテンツモデレーション等の透明性の確保等に関する規律①

参考：第14条「利用規約」

- 第14条「利用規約」は、仲介サービス提供者は、サービス利用規約の中に、サービスの受け手により提供されるコンテンツについて、仲介サービスの利用にあたって利用者に課す条件についての情報を含めるものとするとされている。
- この条件についての情報には、アルゴリズムや人間の判定によるコンテンツモデレーションのために利用される方針、手続き、手段およびツールについての情報を含み、また苦情処理システムにおける処理方法に関する情報を含むものとされている。
- また、その情報については、分かりやすく、ユーザーフレンドリーであること等も求められている。

条文（仮訳）

第14条 利用規約

（第1項）仲介サービスの提供者は、そのサービスの受領者が提供する情報に関して、そのサービスの利用に関連して課す制限に関する情報を、その利用条件に含めなければならない。その情報には、アルゴリズムによる意思決定と人間によるレビューを含む、コンテンツ調整の目的で使用される方針、手続き、手段、ツール、および内部苦情処理システムの手続き規則に関する情報を含めるものとする。その情報は、明確かつ平易で、分かりやすく、ユーザーフレンドリーで、曖昧さのない言語で記載され、容易にアクセス可能で機械可読な形式で一般に公開されなければならない。
（第2項）仲介サービスの提供者は、利用条件に重要な変更があった場合、サービスの受領者に通知しなければならない。
（第3項）仲介サービスが主として未成年者を対象とするものであるか、または未成年者が主として利用するものである場合、当該仲介サービスの提供者は、未成年者が理解できるような方法で、サービスの利用条件および利用制限を説明しなければならない。
（第4項）仲介サービスの提供者は、表現の自由、メディアの自由および多元性、その他本憲章に謳われている基本的権利および自由など、サービスの受け手の基本的権利を含め、関係者全員の権利および正当な利益を十分に考慮した上で、第1項にいう制限の適用および実施において、真摯に、客観的に、かつ、適切な方法で行動しなければならない。
（第5項）超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンのプロバイダーは、サービスの受け手に対し、利用可能な救済および救済メカニズムを含む諸条件の簡潔で容易にアクセスでき、かつ機械が読み取り可能な要約を、明確かつ曖昧さのない言語で提供しなければならない。
（第6項）第33条にいう超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンは、サービスを提供するすべての加盟国の公用語で利用規約を公表しなければならない。

【出典】本WG第5回会合（2024年3月4日）配付資料WG5-1-6「EU DSA法（Digital Services Act）の概観」（株式会社野村総合研究所ご発表資料）

コンテンツモデレーション等の透明性の確保等に関する規律②

参考：第16条「通知と行動の仕組み」、第17条「理由の通知」

- 第16条「通知と行動の仕組み」、第17条「理由の記載」はホスティングサービス提供者に関する追加規定として明記されている。
- 第16条では、ホスティングサービス提供者は、個人または団体が違法コンテンツであると考える特定の情報項目がホスティングサービス上に掲載されていることについて、個人または団体に対して容易かつ電子的に通知できる仕組み（mechanisms）を導入するものとする、とされている。
  - この通知の仕組みは十分に正確で満足できる裏付けのある通知の提出となることを可能にするものでなければならないとされ、違法コンテンツであると考える個人または団体は、その理由の説明や根拠となるURL、通知を行った個人や団体の名前やメールアドレス等の要素が含まれるようにするべきとされている。
- 第17条では、ホスティングサービス提供者が、(a)サービスの受け手（recipients）から投稿された特定のコンテンツを削除あるいはアクセス遮断する、(b)報酬支払いの停止、打ち切りまたは制限、(c)サービス提供の全面的又は一部停止または打ち切り、(d)アカウントの停止または打ち切りにあたっては、これらの措置を講ずるよりも前に、措置を決定したことと、決定に至った理由について、サービスの受け手に説明しなければならない、とされている。

条文（抜粋、仮訳）

第16条 通知と行動の仕組み
（第1項）ホスティングサービスのプロバイダは、個人または団体が違法コンテンツとみなす特定の情報項目がそのサービス上に存在することを通知できるような仕組みを設置しなければならない。これらの仕組みは、アクセスが容易でユーザーフレンドリーでなければならず、電子的手段のみによる通知の提出を認めなければならない。

第17条　理由の通知
（第1項）ホスティングサービスの提供者は、サービスの提供を受ける者が提供する情報が違法なコンテンツであること、またはその利用条件に適合しないことを理由として、以下のいずれかの制限を行う場合、影響を受けるサービスの提供を受ける者に対して、明確かつ具体的な理由を説明するものとする：
(a)コンテンツの削除、コンテンツへのアクセスの無効化、またはコンテンツの降格を含む、サービスの受領者が提供する情報の特定の項目の可視性の制限
(b)金銭の支払いの停止、終了またはその他の制限
(c)本サービスの全部または一部の提供の停止または終了
(d)本サービスの受領者のアカウントの停止または終了

【出典】本WG第5回会合（2024年3月4日）配付資料WG5-1-6「EU DSA法（Digital Services Act）の概観」（株式会社野村総合研究所ご発表資料）

コンテンツモデレーション等の透明性の確保等に関する規律③

参考：第20条「内部苦情処理体制」

- 第20条「内部苦情処理体制」はオンラインプラットフォーム提供者に関する追加規定として明記されている。
- 第20条では、オンラインプラットフォーム提供者は、通知を受けた情報が違法コンテンツまたは利用規約違反であることを理由として下した決定に対して、少なくとも6ヶ月間は通知を提出したサービス受領者に対し、電子的かつ無料で苦情を申し立てることができる効果的な内部苦情処理システムへのアクセスを提供しなければならない、とされている。
  - 6か月の期間は、第16条第5項、第17条に従ってサービス受領者が提出した通知を受けた日が開始日となる。
  - オンラインプラットフォーム提供者は、内部苦情処理システムをユーザーフレンドリーに構築しなければならない。
  - サービス受領者からの通知に対してオンラインプラットフォーム提供者が行う決定には、通知された情報へのアクセスの削除・無効・可視性の制限をする、通知されたユーザーに対するサービスの提供の全部または一部を停止・解約する、通知されたアカウントを停止・解約する、通知されたユーザーが提供した情報を一時停止、終了、またはその他の方法で収益化する能力を制限する、等が挙げられる。

条文（抜粋、仮訳）

第20条 内部苦情処理体制
（第1項）オンライン・プラットフォーム提供者は、通知を提出した個人または団体を含むサービスの受領者に対し、本項に定める決定後少なくとも6ヶ月間は、受領者が提供した情報が違法コンテンツであることまたはその利用条件に適合しないことを理由として、通知を受領した際にオンライン・プラットフォームのプロバイダが行った決定またはオンライン・プラットフォームのプロバイダが行った以下の決定に対して、電子的かつ無料で苦情を申し立てることができる効果的な内部苦情処理システムへのアクセスを提供しなければならない：

(a)情報へのアクセスを削除するか、無効にするか、または可視性を制限するかどうかの決定
(b)受信者に対するサービスの提供の全部または一部を停止または終了するか否かの決定
(c)受信者のアカウントを停止または解約するか否かの決定
(d)受信者によって提供された情報を収益化する能力を停止、終了、またはその他の方法で制限するかどうかの決定

【出典】本WG第5回会合（2024年3月4日）配付資料WG5-1-6「EU DSA法（Digital Services Act）の概観」（株式会社野村総合研究所ご発表資料）

コンテンツモデレーション等の透明性の確保等に関する規律④

参考：第42条「透明性に関する報告義務」

- 第15条で規定された仲介サービス提供者への透明性の報告義務に加えて、第42条「透明性に関する報告義務」は、VLOP・VLOSEに関する追加規定として明記されている。
- 第42条では、VLOP・VLOSEは、遅くとも第33条(6)第2号で言及される申請日から2ヶ月後までに第15条で言及される報告書を公表し、その後は少なくとも6ヶ月ごとに公表しなければならない、とされている。
  - 公表する報告書には、VLOP・VLOSEが、EU内で提供されるサービスに関して、コンテンツモデレーションに充てる人的資源、および当該職員の資格・言語的専門知識、ならびに研修および支援制度、第15条第1項(e)に掲げる正確性の指標及び関連情報を含む必要がある。
  - VLOP・VLOSEは、第37条(4)に基づく各監査報告書の受領後、遅くとも3ヶ月以内に、過度の遅滞なく、各国のデジタルサービスコーディネーターおよび欧州委員会に送信し、一般に公開しなければならない。

条文（仮訳）

第42条 透明性に関する報告義務

（第1項）超大規模オンラインプラットフォームまたは超大規模オンライン検索エンジンの提供者は、遅くとも第33条(6)第2号で言及される申請日から2ヶ月後までに第15条で言及される報告書を公表し、その後は少なくとも6ヶ月ごとに公表しなければならない。

（第2項）超大規模オンラインプラットフォームのプロバイダーが公表する本条第1項の報告書には、第15条および第24条第1項の情報に加え、以下を明記するものとする：
(a)超大規模オンラインプラットフォームのプロバイダーが、連合内で提供されるサービスに関して、加盟国の該当する公用語ごとに区分された、コンテンツのモデレーションに充てる人的資源
(b)(a)で言及された活動を実施する者の資格および言語的専門知識、ならびに当該職員に与えられる研修および支援
(c)第15条第1項第(e)号に掲げる正確性の指標及び関連情報を、加盟国の公用語ごとに区分したもの

（第3項）第24条(2)で言及される情報に加えて、非常に大規模なオンラインプラットフォームまたは非常に大規模なオンライン検索エンジンのプロバイダーは、本条第1項で言及される報告書に、各加盟国のサービスの平均月間受信者に関する情報を含めるものとする。

（第4項）超大規模オンラインプラットフォームまたは超大規模オンライン検索エンジンのプロバイダーは、第37条(4)に基づく各監査報告書の受領後、遅くとも3ヶ月以内に、完了次第、過度の遅滞なく、設置のデジタルサービスコーディネーターおよび欧州委員会に送信し、一般に公開しなければならない。

（第5項）超大規模オンラインプラットフォームの提供者または超大規模オンライン検索エンジンの提供者が、第4項に基づく情報の公表が、当該提供者または当該サービスの受領者の秘密情報の開示につながり、当該サービスのセキュリティに重大な脆弱性をもたらし、公共の安全を損ない、または受領者に危害を及ぼすおそれがあると考える場合、当該提供者は、公に利用可能な報告から当該情報を削除することができる。この場合、プロバイダーは、公開可能な報告書から当該情報を削除する理由を記載した報告書一式を、設立のデジタルサービス・コーディネーターおよび欧州委員会に送付しなければならない。

【出典】本WG第5回会合（2024年3月4日）配付資料WG5-1-6「EU DSA法（Digital Services Act）の概観」（株式会社野村総合研究所ご発表資料）

【参考３】情報伝送PFによるコンテンツモデレーションの実施の促進等に関連する既存の制度の例等（国内）

インターネット・ホットラインセンター（IHC）による送信防止措置の依頼①

● 警察庁の委託を受け、インターネット上の**違法情報**や**自殺誘引等情報**、**重要犯罪密接関連情報**の通報を受理し、ガイドラインに基づいて**警察に情報提供**するとともに、サイト管理者等に**送信防止措置を依頼**。

◆ 対応の対象となる「違法情報」

1. わいせつ電磁的記録記録媒体陳列
2. 児童ポルノ公然陳列
3. 売春目的等の誘引
4. 出会い系サイト規制法違反の禁止誘引行為
5. 薬物犯罪等の実行又は規制薬物の濫用を、公然、あおり、又は唆す行為
6. 規制薬物の広告
7. 指定薬物等である疑いがある物品の広告
8. 危険ドラッグに係る未承認医薬品の広告
9. 預貯金通帳等の譲渡等の勧誘・誘引
10. 携帯電話等の無断有償譲渡等の勧誘・誘引
11. 識別符号の入力を不正に要求する行為（フィッシング行為）
12. 不正アクセス行為を助長する行為
（ID、パスワードの無断掲載）

◆ 対応の対象となる「自殺誘引等情報」

1. 自殺関与
・・・自殺を仄めかしている者に対し、自殺の実行を「手伝う」「請け負う」等と持ちかける情報
2. 自殺の誘引・勧誘
・・・集団自殺の呼びかけ等、他者の自殺を誘引・勧誘する情報

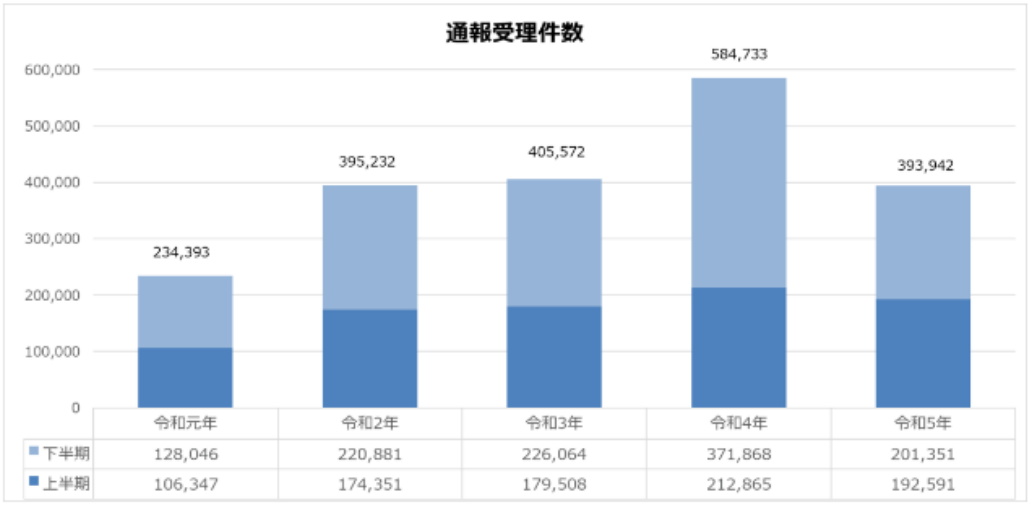
◆ 対応の対象となる「重要犯罪密接関連情報」

1. 拳銃等の譲渡等
・・・拳銃、小銃、機関銃、砲の譲渡等を直接的かつ明示的に誘引等する情報
2. 爆発物・銃砲等の製造
・・・爆発物又は銃砲等の不正な製造を直接的かつ明示的に助長等していると認められる情報
3. 殺人等
・・・殺人、強盗、不同意性交等、放火、誘拐、傷害、逮捕・監禁、脅迫を直接的かつ明示的に請負等する情報
4. 臓器売買・・・臓器売買を直接的かつ明示的に誘引等する情報
5. 人身売買・・・人身売買を直接的かつ明示的に誘引等する情報
6. 硫化水素ガスの製造
・・・硫化水素ガスの製造方法を教示し、その製造を誘引する情報
7. ストーカー行為等
・・・ストーカー行為等の規制等に関する法律のつきまとい等若しくは位置情報無承諾取得等によって不安を覚えさせる行為又はストーカー行為を直接的かつ明示的に請負等する情報
8. 犯罪実行者の募集
・・・具体的な仕事の内容を明らかにせずに著しく高額な報酬の支払を示唆して行う犯罪の実行者を募集する情報
（いわゆる闇バイトの求人・求職）

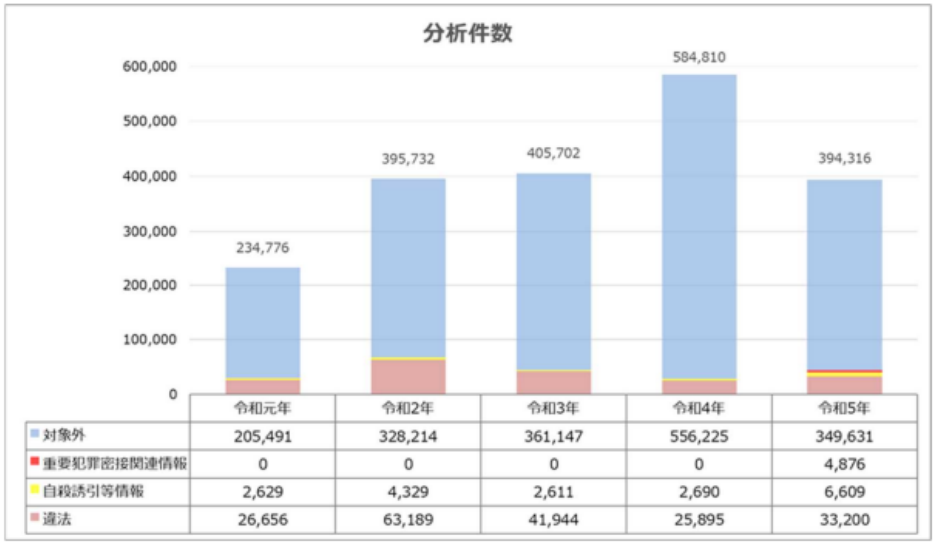
【出典】IHCウェブサイト（https://www.internethotline.jp/）

インターネット・ホットラインセンター（IHC）による送信防止措置の依頼②

- IHCは、2023年1月から12月までの12か月間で**393,942件（運用ガイドラインに基づく分析件数としては394,316件）の通報**を受理。



	令和元年	令和2年	令和3年	令和4年	令和5年
下半期	128,046	220,881	226,064	371,868	201,351
上半期	106,347	174,351	179,508	212,865	192,591



	令和元年	令和2年	令和3年	令和4年	令和5年
対象外	205,491	328,214	361,147	556,225	349,631
重要犯罪密接関連情報	0	0	0	0	4,876
自殺誘引等情報	2,629	4,329	2,611	2,690	6,609
違法	26,656	63,189	41,944	25,895	33,200

◆ 違法情報の処理結果

	分析件数（国 内）	警察へ通報	削除依頼	削除完了
わいせつ電磁的記録記録媒体陳列	2,435 件	2,338 件	1,675 件	1,419 件
児童ポルノ公然陳列	412 件	379 件	199 件	191 件
売春目的等の誘引	27 件	27 件	16 件	16 件
出会い系サイト規制法違反の禁止誘引行為	0 件	0 件	0 件	0 件
薬物犯罪等の実行又は規制薬物の濫用を、公然、あおり、又は唆す行為	2 件	1 件	0 件	0 件
規制薬物の広告	53 件	50 件	22 件	19 件
指定薬物の広告	0 件	0 件	0 件	0 件
指定薬物等である疑いがある物品の広告	0 件	0 件	0 件	0 件
危険ドラッグに係る未承認医薬品の広告	0 件	0 件	0 件	0 件
預貯金通帳等の譲渡等の勧誘・誘引	3 件	3 件	1 件	0 件
携帯電話等の無断有償譲渡等の勧誘・誘引	1 件	1 件	0 件	0 件
識別符号の入力を不正に要求する行為	25 件	19 件	0 件	0 件
不正アクセス行為を助長する行為	0 件	0 件	0 件	0 件
合計	**2,958 件**	**2,818 件**	**1,913 件**	**1,645 件**

※ 削除依頼より5営業日後に確認した際の件数です。令和6年1月末時点では、1,716 件(89.7%)が削除に至りました。

◆ 重要犯罪密接関連情報の処理結果

	分析件数	対応依頼	削除完了
拳銃等の譲渡等	15 件	10 件	8 件
爆発物・銃砲等の製造	16 件	15 件	7 件
殺人、強盗、不同意性交等、放火、誘拐、傷害、逮捕・監禁、脅迫	411 件	356 件	252 件
臓器売買	18 件	16 件	5 件
人身売買	0 件	0 件	0 件
硫化水素ガスの製造	2 件	1 件	1 件
ストーカー行為等	3 件	2 件	2 件
犯罪実行者募集	4,411 件	2,979 件	2,136 件
合計	**4,876 件（4,132 件）**	**3,379 件（3,022 件）**	**2,411 件（2,139 件）**

※ （）内は、サイバーパトロールセンター（警察庁委託事業）からの通報分を内数で示したものです。

※ 削除完了件数は、令和6年1月末に確認した際の件数です。

◆ 自殺誘引等情報の処理結果

	分析件数	対応依頼	削除完了
自殺関与	115 件	115 件	87 件
自殺の誘引・勧誘（集団自殺の呼びかけ等）	6,494 件	6,493 件	3,764 件
合計	**6,609 件（6,530 件）**	**6,608 件（6,529 件）**	**3,851 件（3,804 件）**

※ （）内は、サイバーパトロールセンター（警察庁委託事業）からの通報分を内数で示したものです。

※ 削除完了件数は、令和6年1月末に確認した際の件数です。

【出典】IHCウェブサイト（https://www.internethotline.jp/）

薬機法に基づく厚生労働大臣・都道府県知事による送信防止措置要請

- 薬機法（医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律）は、医薬品等の名称、効能、効果又は性能に関して、**虚偽・誇大広告を禁止**。
- 上記を含めた広告規制に違反する広告について、**厚生労働大臣又は都道府県知事は、特定電気通信役務提供者（情報伝送PFなど）に対し、その送信を防止する措置の実施を要請**できる。
- 上記要請を受けて送信防止措置を講じた特定電気通信役務提供者は、当該措置により発信者に生じた損害について、当該措置が当該広告である情報の不特定の者に対する送信を防止するために**必要な限度において行われたものであるときは免責**。

◆ 薬機法の関連条文

（誇大広告等）

第六十六条　何人も、医薬品、医薬部外品、化粧品、医療機器又は再生医療等製品の名称、製造方法、効能、効果又は性能に関して、明示的であると暗示的であるとを問わず、虚偽又は誇大な記事を広告し、記述し、又は流布してはならない。

2　医薬品、医薬部外品、化粧品、医療機器又は再生医療等製品の効能、効果又は性能について、医師その他の者がこれを保証したものと誤解されるおそれがある記事を広告し、記述し、又は流布することは、前項に該当するものとする。

3　（略）

（違反広告に係る措置命令等）

第七十二条の五　（略）

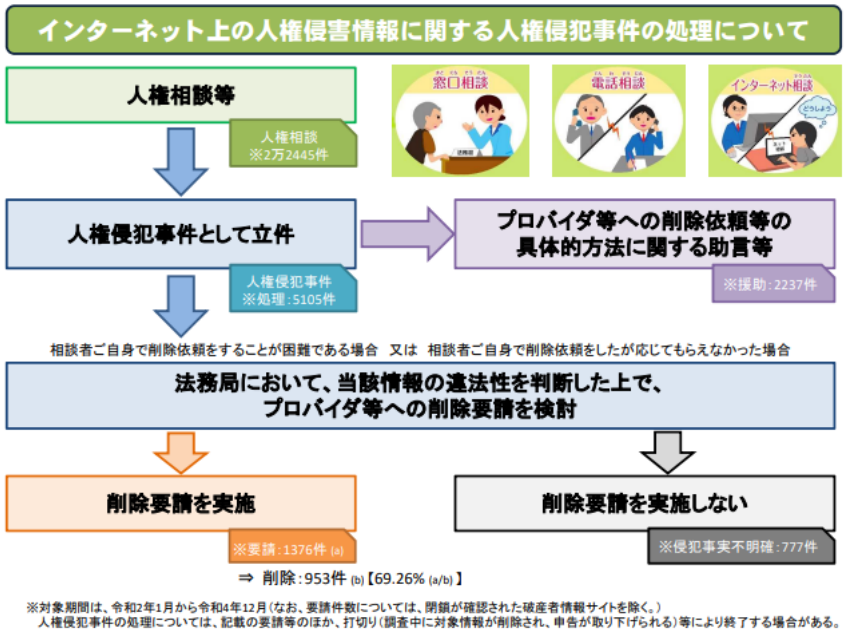
2　厚生労働大臣又は都道府県知事は、第六十六条第一項又は第六十八条の規定に違反する広告（次条において「特定違法広告」という。）である特定電気通信（特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律（平成十三年法律第百三十七号）第二条第一号に規定する特定電気通信をいう。以下同じ。）による情報の送信があるときは、特定電気通信役務提供者（同法第二条第三号に規定する特定電気通信役務提供者をいう。以下同じ。）に対して、当該送信を防止する措置を講ずることを要請することができる。

（損害賠償責任の制限）

第七十二条の六　特定電気通信役務提供者は、前条第二項の規定による要請を受けて特定違法広告である特定電気通信による情報の送信を防止する措置を講じた場合その他の特定違法広告である特定電気通信による情報の送信を防止する措置を講じた場合において、当該措置により送信を防止された情報の発信者（特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律第二条第四号に規定する発信者をいう。以下同じ。）に生じた損害については、当該措置が当該情報の不特定の者に対する送信を防止するために必要な限度において行われたものであるときは、賠償の責めに任じない。

法務省人権擁護機関による人権侵害情報の削除要請

- 名誉毀損、プライバシー侵害、不当な差別的言動等の**インターネット上の人権侵害情報による人権侵犯事件**について、被害者等から相談を受けるなどした法務局は、必要な調査を行い、被害者に対して助言等の援助を行うほか、**プロバイダ等（情報伝送PFなど）に対して任意の削除要請**を行う。



※対象期間は、令和2年1月から令和4年12月（なお、要請件数については、閉鎖が確認された破産者情報サイトを除く。）
人権侵犯事件の処理については、記載の要請等のほか、打切り（調査中に対象情報が削除され、申告が取り下げられる）等により終了する場合がある。

【インターネット上の人権侵害情報】法務省の人権擁護機関による削除要請と削除対応率（サイト別）

番号	サイト名	（種別）	要請件数 ○+△+×−□(件)	削除合計 ○+△（件）			削除せず ×（件）	全部削除率 ○/□（率）	削除対応率 (○+△)/□（率）
					全部削除 ○（件）	一部削除 △（件）			
1	２ちゃんねるブックマーク	掲示板のコピーサイト	20	20	20	0	0	100.00%	100.00%
2	爆サイ.com	掲示板	187	186	182	4	1	97.33%	99.47%
3	FC2	ブログ	26	25	24	1	1	92.31%	96.15%
4	２ｃｈ勢いランキング	掲示板のコピーサイト	17	16	16	0	1	94.12%	94.12%
5	Amebaブログ	ブログ	16	15	15	0	1	93.75%	93.75%
6	ログ速	掲示板のコピーサイト	12	11	11	0	1	91.67%	91.67%
7	ライブドアブログ	ブログ	27	24	22	2	3	81.48%	88.89%
8	YouTube	画像・動画の共有サイト	110	84	83	1	26	75.45%	76.36%
9	Imgur	画像・動画の共有サイト	16	12	12	0	4	75.00%	75.00%
10	Facebook	SNS	13	9	8	1	4	61.54%	69.23%
11	ホストラブ	掲示板	12	8	6	2	4	50.00%	66.67%
12	２ちゃんねる(2ch.sc)	掲示板	85	56	44	12	29	51.76%	65.88%
13	Instagram	SNS	11	7	7	0	4	63.64%	63.64%
14	Yahoo!知恵袋	Q&Aサイト	22	13	9	4	9	40.91%	59.09%
15	５ちゃんねる	掲示板	99	50	44	6	49	44.44%	50.51%
16	みみずん検索	掲示板のコピーサイト	10	3	3	0	7	30.00%	30.00%
17	Twitter	SNS	143	36	31	5	107	21.68%	25.17%
18	2ch2.net	掲示板	12	2	2	0	10	16.67%	16.67%
	その他		538	376	354	22	162	65.80%	69.89%
	全体		1376	953	893	60	423	64.90%	69.26%

※　件数は、個別のプロバイダ等に対する削除要請の件数であり、個別の投稿の件数ではない。通例は、同一の被害者について、特定のサイト等に複数の人権侵害性のある投稿がなされ、そのような複数の投稿について、まとめてプロバイダ等に削除要請を行うところ、このようなプロバイダ等1社に対する要請1回を1件としてカウントしたもの。このうち、全部が削除された場合を「全部削除」、一部が削除されたものを「一部削除」とし、その合計を要請件数で除した数値を削除対応率として示した。
※　削除には、被害者や地方公共団体による削除依頼に基づく削除のほか、投稿者による自主的な削除もある。
※　法務省の人権擁護機関による削除要請と削除との条件関係は、厳密に特定できるものではない。
※　対象期間は、令和2年1月～令和4年12月。対象期間中に処理を終えた要請件数が10件以上のサイト名を掲げた（閉鎖が確認された破産者情報サイトを除く。）。

【出典】総務省「誹謗中傷等の違法・有害情報への対策に関するワーキンググループ（第7回）」（2023年6月1日）配付資料1「インターネット上の人権侵犯事件の処理について」（法務省）

政府が情報伝送PFにコンテンツモデレーションを要求するにあたっての原則（アンケート調査結果）

制限措置に関する原則

政府が、デジタルプラットフォーム事業者（SNS、検索、動画投稿・共有など）に対して、そのオンラインサービス上の偽情報や誤情報を削除等により制限する措置（モデレーション）を要求するにあたって、どのような原則が必要であると思いますか。
あてはまるものをすべてお選びください。（いくつでも）

- 日本の上位3つをみると「法的根拠」（51.4%）、「正当な目的」（51.1%）、「透明性」（44.2%）であった。
- 諸外国の最も高い回答は、日本を除くすべての対象国で「透明性」となった。特に、アメリカ、イギリス、韓国、オーストラリアでは5～6割台となり、日本の4割よりも高い。

	全体	法的根拠（Legal basis）・・・政府による要求が法的枠組みに規定されていること	正当な目的（Legitimate aims）・・・政府による要求が正当な目的に照らし、過剰ではない方法で、かつ、必要性、比例性、合理性等の基準に基づき実施されること	承認（Approvals）・・・政府による要求について、その事前承認の要件（基準や手続等）が、政府による要求の結果として生じる表現の自由等への干渉の程度に見合う法的枠組	透明性（Transparency）・・・政府による要求に関する法的枠組みが明確であり、公衆にとって容易にアクセス可能なものであること（一方、国家安全保障又は法執行の活動への	監督（Oversight）：政府による要求が法的枠組みに基づいていることを確認するために、効果的かつ公平な監督のためのメカニズムが存在すること	救済（Redress）・・・政府による法的枠組みへの違反を特定し、改善するための効果的な司法的・非司法的救済について、投稿等した情報が削除等制限された個人に提供され	その他
日本	(2000)	51.4	51.1	26.8	44.2	29.5	29.6	1.8
アメリカ	(1000)	36.6	44.9	43.9	50.6	29.2	23.1	1.8
イギリス	(1000)	46.9	41.3	42.6	51.6	27.1	24.9	1.6
フランス	(1000)	43.7	35.8	29.7	44.3	26.3	28.1	1.2
韓国	(1000)	60.6	56.0	34.8	62.8	35.8	22.5	0.2
オーストラリア	(1000)	52.5	46.3	47.6	60.9	36.6	29.6	1.5

【出典】検討会第18回（WG第17回）会合（2024年5月9日）配付資料18-2-1「令和5年度国内外における偽・誤情報に関する意識調査 結果紹介」（みずほリサーチ＆テクノロジーズ株式会社ご発表資料）

【参考４】検討会・WGにおけるこれまでの主なご意見等

検討会・WGにおけるこれまでの主なご意見等（論点１関係）

- 偽情報とされるものの中には、意図的に作られた偽情報だけでなく、**悪意はないが間違っている情報**や、**事実関係は間違っていないが異なる文脈で使われることで誤った印象を植え付けるもの**などが含まれており、**その境界はしばしば不明確であり、何が偽情報で何がそうでないかを判別することは容易ではない**。【検討会第1回・森構成員】

- 総務省のこの会の全体的なテーマが、デジタル空間における情報流通の健全性ということになっているが、これをPrinciplesやPracticesというところに当てはめたときに、そもそもどういう状態が健全かの議論とある程度の合意というのをしていくのがPrincipes、原則の議論に相当する。また、健全な状態に持っていくというときに、健全性を誰がどのように担保しているか、また、どのように確認するかということの方法論（Practice）の議論も同時に必要。**健全性、偽情報・誤情報に関しても、何をもって偽情報と誰が判断するのかというのは、おそらく非常に難しいグレーゾーンを含んでいるようなものもあるかと思う**が、このような点に関し、広くAIガバナンスという観点も含め、実際にはこの（PrincipleとPracticeの）両方が行き来をしたりするということが重要。【検討会第3回・江間構成員】

- 各サービスに共通している考え方としては、**信頼できる機関によるファクトチェック結果に基づいて明らかに偽・誤情報と判断できるようなもの**については、対応するということにしている。具体的な禁止行為はサービスの性質によって若干異なっているが、特に自由に議題が設定できるような旧LINEのサービス、オープンチャットとかVOOMに関しては、拡散という行為に着目して禁止事項を設定している。【検討会第9回／WG第3回・LINEヤフー】

- プラットフォーム事業者としては、伝統的に真偽判断の問題について一定の立場を取ってきたという中で、**いかにしてその問題が偽情報の問題として対応すべきかどうかということを判定して、その場合にそうした真偽判断に関する原則的な立場をどういうふうに修正していくのかという点が問題**になってくるだろう。【検討会第9回／WG第3回・LINEヤフー】

- プラットフォーム事業者としては、**情報制約がある中で、ある意味で見切り発車が求められているというようなところもある**かと思う。**投稿削除を行った場合に、蓋を開けてみると実は事後的に見るとリスクはなかったというようなケースもあり得るわけですけれども、そうした対応を行っていくことが、これまで例えば名誉毀損事案なんかにおける伝統的なプラットフォームとしての立場と比較して、真偽判断の在り方としてよいのだろうかということ**は、常に悩みを抱えながら日々業務をやっているところ。【検討会第9回／WG第3回・LINEヤフー】

- **真偽は白黒がはっきりしているものではない**。部分的にでも偽の情報がある際に、それを削除しなければいけないということになるか？ということにもなる。そもそも**一民間の企業にとって、それが本当に真であるか、あるいは偽の情報なのかということを決めることはできない**。もちろん民間であれ公共であれ、ある一つの組織がそれを裁定することができるものが存在するとも考えられない。海外の国において見られる例でも見られますように、それがオーバーリーチとなってしまって、逆に情報の統制という形で表れる場合もある。またさらに、**フェイスブックやインスタグラム上にあるものは全て真実であるべきという想定に基づいてしまうと、例えばユーモアや風刺なものも含めて、それが排除されることになってしまう**。【検討会第14回／WG第10回・Meta Platforms Inc.】

- やっぱりDSAが結局偽情報に対してどういうスタンスを取っているのか、あるいは考え方というんですか、思想というんですか、そういうものがやっぱりちょっと分かりにくいなというのが率直なところで、結局例えば**偽情報と違法コンテンツというのが例えば完全にセパレートされた概念なのか、あるいは偽情報も場合によっては違法コンテンツに含まれるのか**という。どうも前文を見ると、**偽情報というのは、違法コンテンツの文脈とは異なる、自主規制の文脈**だみたいなことが書いてあって、偽情報と違法コンテンツとの関係性というところも少し見えにくいところがあったかなと思った。【WG第5回・山本(龍)主査】

検討会・WGにおけるこれまでの主なご意見等（論点２関係）①

- **発信者側の情報、特にAIを使ったかが一番大きいと思うが、そういう部分が、なるべく分かりやすく情報開示されている状態になった方が多分判別率が上がることがある**かとも思う。**プラットフォームの側も、例えば個人が情報発信するに当たっても、できる範囲でスクリーニングして、そういう注記を出していくこと**を対応として求めていくことが、偽・誤情報の拡散防止という意味では比較的効果がありそう。権利との調整は最終的に必要だが、必要性、有効性という意味では意味がありそう。【検討会第2回・落合構成員】

- **DSAは、コンテンツモデレーションというものに対して恐らく初めて明確な法的定義を与えていて**、まずは違法情報の削除、あるいは利用規約に適合しない情報、有害情報や、ディスインフォメーションが広く含まれる。しかし、**その定義はあくまでデジタルプラットフォームの側がする形**になっている。識別したり、**削除のほかにも非常にいろんな手段**がある。**降格**、**収益不能化・デマネタイゼーション**、**アクセス不能化**、**可視性に影響を与える措置**など、非常に広くとっており、**例えばしばしば話題になるようなシャドーバンのようなものも含まれてはくる**。【検討会第4回・生貝構成員】

- 民主主義の維持・促進のためには信頼可能な情報に基づく対話的コミュニケーションが確保されることが必要であり、当然そのベースラインとなる信頼可能な情報がしっかり流通していないといけないということになるんだろうと思う。基本的な方向性としては、メディアの持続可能性、特にローカルのテレビ局とか新聞各社持続可能性が重要になってくるだろうと思うし、基本情報の作成・流通、それから**信頼可能な情報をプロミネントなものにしていくこと、特に放送コンテンツ等のプロミネンスも、そういう意味では重要になってくる**のかなと考えている。【検討会第6回・山本(龍)座長代理】

- 情報空間のインターネット空間の中でも、欧米などで、**プラットフォームやインターネットサービス事業者が提供するコンテンツ自体について、放送に準じたような何らかの対策を求めていくこと自体も、プロミネンスとかモデレーションを超えて、そういう対応を求めていく場合もある**かと思う。【検討会第6回・落合構成員】

- 私も**コンテンツ・モデレーションは、大きく2つに分類できる**と思っており、**よくないコメントとかを消すとか、見えなくするという方法**もあるし、何で評価するかは別にして、**よいコメントを目立つ位置に、見やすい位置に置くという方法**があり、**これを併合してやっていただくというのが実は偽情報対策にとっても非常に重要**なのではないかなと考えている。特にAIに関しては、こうやってプラットフォーム間で協力があるといいのではないかなと思っていた。【検討会第10回／WG第4回・水谷構成員】

- **「コミュニティノート」を見てみると、一般ユーザーの参加によって災害時に特に即応的に必要なときに偽誤情報関連の修正情報の発信とか注意喚起で一定程度の役割を果たした可能性というのもある**と思っていて、この辺、もう少し慎重に見ていく必要があると思っている。ただし、「コミュニティノート」の脆弱性とか限界というのは、海外の例えばウクライナ、それからガザなどの事象においてもそれぞれ指摘がされているところで、**例えば外部グループによる組織的な操作に弱いとか、「コミュニティノート」自身が偽誤情報の流通に寄与するとか、それ以外にも、ボトムアップ的なものなので即応的に対応できる一方、政治的なものに寄ってしまうなどのバイアスが指摘**。この辺りが日本語の「コミュニティノート」でどうなのか、また、偽誤情報対応のところでどう役に立っていくのかというところも慎重な議論が必要。【検討会第10回／WG第4回・澁谷構成員】

- やはり一つや少数の対応策での包括的な対応というのは非常に難しいんだろうなと思っていて、**「コミュニティノート」や、それからファクトチェック団体、それから政府や有識者、あるいは専門家による発信や、あるいはもうちょっと別のモデレーション的なもの、そして人のチェックによる削除とか、いろいろな多面的な取組を推し進めるというのが１つの方向性**とは感じている。【検討会第10回／WG第4回・澁谷構成員】

検討会・WGにおけるこれまでの主なご意見等（論点２関係）②

- 私どものポリシーの第2の柱が、コンテンツ削除のポリシーに違反はしていないけれども、そういった**ミスインフォメーションのコンテンツの配信自体を減らす、もしくは降格させる**こと。私どもが降格するコンテンツの中で最も広く知られているのは、私どもに対して**第三者のサードパーティーのファクトチェッカーによって評価されたポスト、投稿**。この当社のサードパーティーファクトチェックプログラムは業界でも最大の規模であり、そして大きな投資をしている。【検討会第14回／WG第10回・Meta Platforms Inc.】

- 第3の柱が、人々に対して情報を知らせるという役割。**ミスインフォメーションであるというものに関して、それをきちんとユーザーに対して情報として通知をする**、そしてまた、それが繰り返し行っているものに対しての通知。こういった**コンテンツに対する警告画面ですとかラベル**によりまして、驚異と評価されたコンテンツを繰り返し投稿する利用者の投稿に対して、そういった警告のプロンプトを表示することによって、それを見る利用者自身が注意を喚起することができる。例えば、ファクトチェッカーによって「虚偽」と評価されたコンテンツを繰り返し投稿している利用者も対象となるし、あるいは**既にそれは正しいものではないとされた投稿をさらにシェアした利用者への通知**ですとか、あるいは**ファクトチェッカーによって論破されたコンテンツに対する警告画面やラベル**が含まれる。【検討会第14回／WG第10回・Meta Platforms Inc.】

- TikTokのモデレーションは全てコミュニティガイドラインを基準に行っているところ、特に偽・誤情報についても、明確に削除の対象となる基準を定めている。2種類あって、まず1つ目が禁止されるもの。これは、発信者の意図に関わらず、個人や社会に重大な危害を及ぼし得る虚偽の情報ということで、仮に善意に基づくものであったとしても、有害な偽・誤情報は禁止。次に2つ目だが、今度は、**おすすめフィードの対象外になるものとして、一般的な陰謀論や緊急事態に関連する未確認の情報が含まれるコンテンツなどが含まれている**。【検討会第14回／WG第10回・TikTok Japan】

- **コミュニティノートについて、効果自体については、具体的に検証できつつある**と考えており、**フォローアップ自体が35%、あるいは、具体的なシェアであったり、こういったものが大幅に低下するというようなところも見てとれている**ので、この部分については、さらに推移を見ていきたいと思うが、効果は基本的には定量的にも表れているという認識。【検討会第15回／WG第11回・X(Twitter Japan)】

- 透明性に関しては、**AI利用の表示を求めさせること**もすでに各所で提案されており、ここでも論点となるのではないか。【WG第16回・曽我部主査代理 ※会合後の追加意見】

検討会・WGにおけるこれまでの主なご意見等（論点５関係）

- 偽情報とされるものの中には、意図的に作られた偽情報だけでなく、悪意はないが間違っている情報や、事実関係は間違っていないが異なる文脈で使われることで誤った印象を植え付けるものなどが含まれており、その境界はしばしば不明確であり、何が偽情報で何がそうでないかを判別することは容易ではない。そのような状況で、**①何を削除するか、②どのくらいの数を削除するか、といったことについて法制度を作ったり、統一的な基準を設けたりすることは必ずしも適当ではない**。プラットフォーム事業者に過度の削除圧力をかけることは、当該プラットフォームに情報を投稿する利用者の表現の自由とプラットフォーム事業者自身の表現の自由を共に脅かすことにつながる。【検討会第1回・森構成員】

- **偽情報の流通に利用されるプラットフォーム事業者は、コンテンツモデレーション等の偽情報対策を実施することについて、社会からの強い期待**を受けている。【検討会第1回・森構成員】

- アメリカでも今問題になっているが、政府が権力を持って表現空間に介入してくるというのがどこまで許容されるべきか。**政府と事業者の間の透明性とかアカウンタビリティの向上も同時に確保されていくべき**。【検討会第1回・水谷構成員】

- アメリカの議論で参考になるのは、DPF事業者を検閲の代理人化させてはいけないので、**事業者への政府機関のコンテンツ削除要請なんかに関しては、政府機関側の、政府機関サイドの透明性をいかに確保していくか**が今後、非常に重要になってくる。【検討会第4回・水谷構成員】

- 政府側の透明性を確保する方法としては、もちろん色々と方法はあると思うが、一つは**法律で、こういう部分についてはきちんと透明性レポートを政府が逆に出すというのを定めるということ**が一つ。例えば、アメリカで提案があったソーシャルメディア検閲透明化法を参考にすると、政府の職員が表現の自由とか、そうしたことに関係しそうな案件でDPF事業者と接触した場合は、それをきちんと記録として残せと。それを事後的にチェックできるようにしろ、というようなことを規律で定める。そこまでできるかどうかは別にして今、DPF事業者側が政府からの要請についての記録を透明性レポートで出してくれている。だから、あれに対応するものを逆に政府側が、例えば白書とかできちんと出すだけでもすごく変わってくるので、**個々のコンテンツの削除について、どういうものをやっているのかということをきちんと明示していただくための仕組みというのを、何かしら作るというのが重要**。【検討会第4回・水谷構成員】

- アメリカの議論の拝見をしていて、いろいろな主体がそれぞれ透明性を持っていくことについて、分け隔てなく、何らかのアクションをするものについては、それぞれ透明性を持っていくというのが一つ、意味がある対応なのではないかという示唆もあると受け止めた。日本の中でもそういう意味では**情報発信者とか、情報の媒介者、または、それに対して影響を与え得る政府と、それぞれの主体が透明性を高めていくことも一つ大事**なのではないかとも思う。【検討会第4回・落合構成員】

- 透明性だけではなく、**モデレーション等そのものの規律について、もう少し踏み込むことはできないか**。例えば、信頼できるメディアやファクトチェック機関による記事の優先表示、コンテンツモデレーションポリシーの選択権、コミュニティノート、本人認証機能の実装のような工夫を求めることなど。【WG第16回・曽我部主査代理　※会合後の追加意見】

検討会・WGにおけるこれまでの主なご意見等（論点６関係）①

- EUのデジタルサービス法（DSA）のように、**コンテンツモデレーションのポリシーの公表や、モデレータに実施している訓練内容や、AIによる自動処理のエラー率などの記載を求めていくこと**も一案。また、**削除やアカウント停止などの対象になったユーザーに具体的に理由を説明すること**や、**判断が間違っていた場合の対応など苦情処理体制の整備**も求めていく必要。【検討会第1回・森構成員】
- **重要なのは透明性の確保。目指すべき社会をしっかりと考えて、具体的な透明性・アカウンタビリティーの確保を促していく、プラットフォーム事業者に促していくということがとても重要**。例えば、どういうデータを公開する必要があるとかということをきっちりと定義づけし、その定義の果てに得られた結果が、しっかりエビデンスベースで有効な対策を検討していくことができるという状態にしておく。これが社会としては重要。【検討会第2回・山口構成員】
- 課題として思っているのが、**具体的に何をどういうふうに透明性を確保して、それをどういうふうに活用するかというところの具体**を詰められていないんじゃないか、並びに、**それを外資系の企業も含めてどのように実行していくか**。さらに、**日本ローカルの透明性をどのように持たせていくか**。あるいは**ユーザーに日本語で対応できる体制をつくる、こういったことを求めていくこと**が大事。【検討会第2回・山口構成員】
- DPF規制について、EUはむしろ、コンテンツの管理をもっとやれ、デジタルサービス法でもシステミックリスクをちゃんと分析評価して、軽減措置をちゃんととりなさいというような規制が入っていたりするが、それとアメリカの規制の議論は全く異なる側面がある。一方で**DPF事業者の透明性、不透明さ、彼らのコンテンツの管理のプロセスの不透明さをめぐる議論というのは、実は世界的に共通した話題としてアメリカやそれ以外でもあがったりしている**。【検討会第4回・水谷構成員】
- **透明性というのはあくまで目的ではなくて手段**だと考えている。ではそれで何を達成するのかという点で、一つは**DPF事業者がユーザーや市民社会に対するlegitimacyを醸成する役割**や、あるいは、有名な裁判官の言葉だが、日光は最高の消毒液であるというようなことが情報公開とかの文脈でよく言われるため、**透明性を高めることによって内々で行われている不誠実な対応に対する抑止効果が出るだろうというようなこと**が考えられる。また、**規制を何かしら入れた場合の効果測定のため、モニタリングするため**にも透明性は要るし、あるいは、**AI等の技術利用がどういうリスクをもたらすかというのも実は分からないところがあるので、これもモニタリングするためには透明性が重要**になってくる。【検討会第4回・水谷構成員】
- **コンテンツモデレーションの仕組みはだいぶシステミックな仕組みになっているので、個別のエラーについて事後的に責任追求するということが、非常に難しい**。そのうえで、**DPF事業者の場合は、コンテンツモデレーションの部門と経営部門というのが基本的には報道機関のように分離していない**、ピラミッド型でモデレーションの部門、トラスト＆セーフティ部門等がある。つまり報道機関で言われるような編集と経営の分離といったエシカルな議論を基本的には採っていない。しかも、モデレーション実務の多くを外注しているといった側面もあり、これがモデレーションの責任の所在を外からますますわかりにくくさせているで、**責任の所在を議論するためにもコンテンツモデレーションのルール形成とか、コンテンツ管理のオペレーションに関する政策といった点に関する年次計画みたいなものをまずは示してもらう必要がある**のではないか。そこにどういう人たちが関わっていて、どのようなプロセスでできているのかという、ここも結局、透明性の話になるが、責任所在をはっきりするためにも透明性の確保が重要になってくる。【検討会第4回・水谷構成員】
- （DSAでは）**利用規約にこういうコンテンツモデレーションをどうやっているのか、どんな人員体制でどんなモデレーションをアルゴリズムを使ってやっているのかといったことを記載**する。特に大きいこれらの方々は**各加盟国、少なくとも27プラス、EEAの言語で提供**しなければならない。そして、それをどのように実施したかということについて、**少なくとも年に1回、VLOPは年に2回、透明性レポートというのを提示**する。政府の透明性という意味だと**各国政府からどんなコンテンツ削除の要請があったり、情報請求の要請があったのか**、あるいは**違法・規約違反別の対応件数**、それから**大体対応1件あたりに中央値としてどのぐらいかかったのか**。それから**コンテンツモデレーションの人員では、大体何語ができる人がどのくらいいて、どんなトレーニングをしているのか**。または、**自動処理のエラー率指標と、それからそれのエラーがあった時にどういうセーフガードをとっているのかといったようなこと**を事細かに出さないといけない。【検討会第4回・生貝構成員】

検討会・WGにおけるこれまでの主なご意見等（論点６関係）②

- **ディスインフォメーション対策等で重要なのは、何語ができるモデレーターをどのくらい抱えているか**。Xは、言語の偏りが大きいといったようなことも言われているところだが、例えばオランダ語ができる人は1人確保しているだとか、イタリア語ができる人は2人確保しているだとか、（DSAでは）そういったことがしっかりと透明に公開をされることになっている。【検討会第4回・生貝構成員】
- （DSAにおける）透明性の側面として面白いのは、コンテンツが削除されたりする。それはしばしば、大体自動的にやるので間違うこともあるといった時に、**コンテンツのモデレーションをやったら、その影響を受けた情報発信者等にちゃんと一々、1件1件理由の通知**をしないといけないことになっている。そしてその通知を受ける、そうすると初めて反論の確保、機会が確保される。GDPRのプロファイリングのアナロジーで言えば、まさに**機械の決定だけに服さずに人間の関与を求める権利**と言っても良い。こういった苦情があった時は、これには別のルートで個人で削除してくれと言ったのに、例えば誹謗中傷等、削除してくれないといったようなものの苦情処理も、ちゃんと理由を通知した上でこういうシステムの対象にしないといけないが、ちゃんと公平な判断をしないといけない。**それでも納得いかなかったら、例えばADRを使うことができる**といったようなことを定めていたりする。【検討会第4回・生貝構成員】
- 利用規約が大事であるがコロコロ変わる、フォローできなければ後で確認可能性がないということで、**利用規約が変わっていくのをちゃんと時系列でもしっかりと把握した上で、変遷も後でチェックできるようなデータベース**も作っている。こういうテクノロジーに基づく、ある種の透明性の確保というのもいろんな形で重要なんだろうなと思う。【検討会第4回・生貝構成員】
- モデレーションの透明性と救済というのは**果たしてどのような情報空間が健全であり、望ましいのかということを社会で全体で議論**する、**それに対して個人が異議を申し立てる**、そういうのを大前提と考えるべきであろう。【検討会第4回・生貝構成員】
- （AIを用いたコンテンツモデレーションに関し）透明性と一口にいっているけれども、どこまで開示すれば十分といえるのかというのは、論文とか読んでいてもいろんなレベルで議論されているところがある。あくまで私個人が考えているところだが、**特にAIがどの程度、プラットフォームのモデレーションに食い込んでいるのかというのをきちんともっと明示していただく必要**。恐らくコンテンツの種類によってAIが適切に使える、使えないといった差があり、例えば、難しいのはヘイトスピーチだと思う。ヘイトスピーチは国によって文脈が全然違い、ある国ではヘイトスピーチになるものが、ある国では全然ヘイトスピーチとして通じないというようなことがあるので、そこの問題がおそらくあるため、まずは、どこまでAIが入っているのかというところ。もう一つは、絶対エラーは一定程度出ているはずなので、このエラー率、どれぐらいエラーが出てしまっているのか。**AIを入れたけど、特定の分野ではむしろエラー率が高いといったことが分かれば、どこに課題があるかわかるので、偽陰性、偽陽性がどの程度、出てきているのかというのをきちんと開示していただくこと**が、ポイントではないか。【検討会第4回・水谷構成員】
- まず、これはアメリカに限らないというか、アメリカのプラットフォーム事業者は御存じのとおり、世界でビジネスを行っていて、日本にも影響を与えているので、世界共通の現象かと思うが、プラットフォーム事業者によるコンテンツモデレーションの大まかな枠組みを見ていくと、まず、利用規約やポリシーにより、コンテンツモデレーションの方針、例えば、投稿が禁止されるコンテンツの内容であるとか、削除基準等について定めた上で、こうした方針に基づいて、利用規約やポリシーに違反するコンテンツを削除したり、当該コンテンツを非表示にしたり、拡散を抑制したり、違反を繰り返すユーザーのアカウントを停止するなどの措置が行われてきた。こういったことは以前から行われてきたが、**特に2021年1月に、ツイッターやフェイスブックが連邦議会議事堂襲撃事件後に、当時のトランプ大統領のアカウントを停止したことで、世界的に大きな衝撃を与え、プラットフォーム事業者によるコンテンツモデレーションの在り方が大きな議論となってきた**。その中で、プラットフォーム事業者によるコンテンツモデレーションの在り方について、公平性や透明性の観点から批判も起きてきたわけだが、**こういった批判も踏まえて、主要なプラットフォーム事業者は、コンテンツモデレーションのアカウンタビリティや透明性を高めるためのガバナンスの構築を進めてきた**。代表的な取組の一つとして、フェイスブックの監督委員会などが挙げられることが多いのかと思う。【WG第8回・成原准教授】

検討会・WGにおけるこれまでの主なご意見等（論点７・８関係）

- 透明性だけではなく、モデレーション等そのものの規律について、もう少し踏み込むことはできないか。例えば、信頼できるメディアやファクトチェック機関による記事の優先表示、コンテンツモデレーションポリシーの選択権、コミュニティノート、**本人認証機能の実装のような工夫を求めること**など。【WG第16回・曽我部主査代理　※会合後の追加意見】

- デジタルサービス法は、まず、媒介サービスを**導管**、**キャッシング**、**ホスティング**と3つに電子商取引指令と同じように分けた上で、その中で**コンテンツを預かるサービスに追加的な義務**が課されて、その中で特に**コンテンツを配布するようなサービスはOPという形でまた義務が増えてきて**、**一番大きいものは4,500万人以上のユーザー数、EU域内の10%以上の人口**といったところで、御紹介するシステミックリスクの評価と軽減といったようなことの義務等々がかかってくる。他にも取引OPに関わる特別の規律なんかもあったりするが、VLOSEは、ほぼVLOPと同様の義務がかかる。【検討会第4回・生貝構成員】